

# English natural sentences automatic identification and word querying free automatic processing method

Publication number: CN1266238

Publication date: 2000-09-13

Inventor: LIN XINGUANG (CN); SONG JIANFU (CN); ZHU MING (CN)

Applicant: YINGYEDA CO LTD (CN)

Classification:


- international: G06F17/28; G06F17/28; (IPC1-7): G06F17/28

- European:

Application number: CN19991002486 19990304

Priority number(s): CN19991002486 19990304

Also published as:

 CN1154066C (C)

[Report a data error here](#)

## Abstract of CN1266238

The present invention relates to a method for automatically identifying English natural sentence and automatic processing method having no need of seeking words, and is characterized by that it utilizes the data base which is undergone system analysis in advance and in which several difference levels of words and phrases are respectively stored to make user reading an English article can automatically pick up all words and phrases of correspondent level in the article and make translation so as to help the user to read English article. Besides, according to different English degrees said data base can freely regulate hard and eas level of all words and phrases, and when the partial words of some level are become simpler, it can make user arbitrarily reduce level of all words.

-----  
Data supplied from the [esp@cenet](mailto:esp@cenet) database - Worldwide

[19]中华人民共和国国家知识产权局

[51]Int. Cl<sup>7</sup>

G06F 17/28

## [12] 发明专利申请公开说明书

[21] 申请号 99102486.9

[43]公开日 2000年9月13日

[11]公开号 CN 1266238A

[22]申请日 1999.3.4 [21]申请号 99102486.9

[71]申请人 英业达股份有限公司

地址 台湾省台北市

[72]发明人 林信光 宋建福 朱 明

[74]专利代理机构 柳沈知识产权律师事务所

代理人 吕晓章

权利要求书 3 页 说明书 4 页 附图页数 10 页

[54]发明名称 英文自然句自动识别及免查单词的自动处理方法

[57]摘要

一种英文自然句自动识别及免查单词的自动处理方法,利用事先经过系统分析,分别存储有数个不同级别单词及短语的数据库,使使用者浏览一篇英文文章时,即可自动根据使用者选定的级别,拾出文章中所有相对应级别的单词及短语,加以翻译,令使用者可阅读英文,该数据库可针对不同英文程度,自由调整所有单词及短语的难、易级别,当某一级别的部分词汇,已变成较简单时,可令使用者任意地降低各词汇的级别。

ISSN 1008-4274

## 权 利 要 求 书

1. 一种英文自然句自动识别及免查单词的自动处理方法, 其特征在于, 该方法是:

5       (一) 先将数据库读入, 以建立相对应的索引编号、级别及具体信息的地址及长度的索引关系, 并将不同级别的单词及短语, 放入其相对应的存储空间内;

          (二) 再输入一篇文章并设定当前的搜索单词序号初始化;

          (三) 计算文章中有几句、几个单词;

10       (四) 再判断计算的单词个数是否大于数据库词汇总数, 如是, 则结束, 如否, 则继续第五步骤;

          (五) 再判断句数标记是否小于数据库内总句数, 如否, 则结束, 如是, 则进行第六步骤;

15       (六) 通过一标志量, 用以区分当前所处理的句子是否在同一显示屏上, 如是, 则置标志量为在同一显示屏上, 并继续进行显示, 且回到第四步骤而重复上述的步骤, 如否, 则继续第七步骤;

          (七) 将当前的显示屏刷新, 重新计算当前的显示位置并显示其相对应的内容, 同时, 将搜索单词序号加 1 并置标志量为不在同一显示屏上, 且回到第四步骤而重复上述的步骤。

20       2. 如权利要求 1 所述英文自然句自动识别及免查单词的自动处理方法, 其特征在于, 该数据库主要包括下列四个地址, 以将词汇划分成数个等级, 并进行整理及排序:

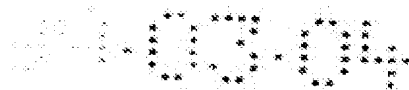
          一索引编号地址, 该索引编号地址用以存放每一单词或短语的索引编号;

25       一单词或短语地址, 该单词或短语地址用以存放数个与索引编号地址相对应的单词或短语;

          一级别地址, 该级别地址用以存放数个与单词或短语地址相对应的级别;

30       一具体信息的地址及长度地址, 该具体信息的地址及长度地址用以存放数个与单词或短语地址相对应的单词或短语的位置及长度。

3. 如权利要求 1 所述英文自然句自动识别及免查单词的自动处理方



法，其特征在于，进行第三步骤中包括如下步骤进行单词查找处理：

(1) 根据使用者设定的单词级别，开始搜索显示器上待阅览文章的词汇；

(2) 再判断是否已搜索到文章尾，如是，则结束单词查找处理，如否，  
5 则进行第三步骤；

(3) 将该篇文章根据自然语言段落规则，分割成数个独立的句子，并于  
总句数加1；

(4) 再判断是否已搜索到句尾，如是，进入第二步骤操作，如否，则进  
行第五步骤；

10 (5) 依规则取出其中的单词，过滤重复的单词，并将总句数加1，再重  
复第四步骤操作。

4. 如权利要求1所述英文自然句自动识别及免查单词的自动处理方  
法，其特征在于，进行第六步骤中包括如下步骤进行单词翻译、显示处理：

15 (1) 根据所搜索到的词汇，计算该词汇在显示器上所对应的位置及长  
度；

(2) 重新调整该词汇对应的行与相邻行之间距离；

(3) 通过数据库具体信息的地址及长度，以及该词汇所对应的位置，计  
算出该词汇的中文翻译存放的位置与长度；

20 (4) 将该词汇所对应的中文翻译正确地放在相邻该词汇的相对应位置  
上；

(5) 于该中文翻译下方加一下划线。

5. 如权利要求1所述英文自然句自动识别及免查单词的自动处理方  
法，其特征在于，进行第七步骤中包括如下步骤进行单词翻译、显示处理：

25 (1) 根据所搜索到的词汇，计算该词汇在显示器上所对应的位置及长  
度；

(2) 重新调整该词汇所对应的行与相邻行之间距离；

(3) 通过数据库的具体信息的地址及长度，及藉该词汇所对应的位置，  
算出该词汇的中文翻译存放的位置与长度；

(4) 将该词汇所对应的中文翻译正确放在相邻该词汇的相对应位置上；

30 (5) 于该中文翻译下方加一下划线。

6. 如权利要求4或5所述英文自然句自动识别及免查单词的自动处理方



法，其特征在于，重新调整该词汇所对应的行与相邻行之间距离可为一般行与行之间距离的两倍。

- 5 7. 如权利要求2、3、4或5所述英文自然句自动识别及免查单词的自动处理方法，其特征在于，该数据库并提供一界面，使针对使用者不同英文程度，可自由调整所有单词及短语难、易的级别，且令使用者可将不在数据库中的词汇，加入数据库中的相对应级别中，或将欲删除的词汇，从数据库的相对应级别中删除。

英文自然句自动识别及免查  
单词的自动处理方法

5

本发明涉及一种英文自然句自动识别及免查单词的自动处理方法，其是指一种能根据使用者选定的级别，自动对所阅览的文章内所有相对应级别的单词及短语进行翻译，使使用者能更方便地阅读及学习英文的自动处理方法。

10

众所周知，英文是国际交流的主要语言工具，能够方便、快速地阅读及理解各类英文文章，是许多人的需求，而提供一最为实用、方便的辅助方式，则为许多语言工具所欲达到的目标，通过电脑并配合适当的软件，以大幅度地提高该各语言工具对语言学习及交流上的方便性。

15

但是，一般英文辅助类学习工具及翻译软件，仅能针对文章中单一单词进行翻译，即游标指到的单词(如图 10 所示)，或整篇文章一次翻译，而无法实际根据使用者的英文程度，翻译出文章中使用者所无法辨识的单词，因而造成使用者须浪费时间一次一个单词查询，或从整篇翻译文章中找寻不熟悉的单词，故并无法满足轻松阅读英文的目标。

20

鉴于此，发明人是针对以上所述的缺陷及依据多年来从事制造的电脑软件产品相关经验，经细心观察研究，终于开发出本发明的一种英文自然句的自动识别及免查单词的自动处理方法。

25

本发明的目的在于提供一种英文自然句自动识别及免查单词的自动处理方法，该方法是利用一事先经过系统分析，而分别存储有数个不同级别单词及短语的数据库，令当使用者浏览一篇英文文章时，可自动根据使用者选定的级别，而抓出文章中所有相对应级别的单词及短语，并加以翻译，以令使用者可依本身的英文程度，而轻松地阅读及学习英文。

本发明的另一目的在于该方法在数据库中提供一可自动记录所有变更的功能，即当变更数据库的内容时，该数据库会重新进行分析与处理，并通过变更的内容更新数据库中的数据。

30

本发明的目的是这样实现的，即提供一种英文自然句自动识别及免查单词的自动处理方法，该方法是：



(一) 先将数据库读入, 以建立相对应的索引编号、级别及具体信息的地址及长度的索引关系, 并将不同级别的单词及短语, 放入其相对应的存储空间内;

(二) 再输入一篇文章并设定当前的搜索单词序号初始化;

5 (三) 计算文章中有几句、几个单词;

(四) 再判断计算的单词个数是否大于数据库词汇节总数, 如是, 则结束, 如否, 则继续第五步骤;

(五) 再判断句数标记是否小于数据库内总句数, 如否, 则结束, 如是, 则进行第六步骤;

10 (六) 通过一标志量, 用以区分当前所处理的句子是否在同一显示屏上, 如是, 则置标志量为在同一显示屏上, 并继续进行显示, 且回到第四步骤而重复上述的步骤, 如否, 则继续第七步骤;

(七) 将当前的显示屏刷新, 重新计算当前的显示位置并显示其相对应的内容, 同时, 将搜索单词序号加 1 并置标志量为不在同一显示屏上, 且回到  
15 第四步骤而重复上述的步骤。

下面结合附图, 详细说明本发明的实施例, 其中:

图 1 为本发明的数据库的结构示意图;

图 2 为本发明的数据库结构实施例的示意图;

图 3 为本发明的硬件结构示意图;

20 图 4 为本发明的主控制程序的流程图;

图 5 为本发明的文章剖析及统计的流程图;

图 6 为本发明的免查词汇的拾取及显示的流程图;

图 7 为本发明的显示词汇翻译界面的示意图;

图 8 为本发明的数据库修改的流程图;

25 图 9 为本发明的数据库修改界面的示意图;

图 10 为现有的显示词汇翻译界面的示意图。

请参阅图 1 和图 2 所示, 本发明为一种英文自然句自动识别及免查单词的自动处理方法, 本发明事先经过一系列的系统分析, 将所有单词及短语依其难、易程度的不同, 划分成不同的等级, 并以此建立一数据库, 该数据库  
30 在将该各词汇划分成数个等级时, 该数据库主要包括四个地址: 一索引编号 ( 101 ) 地址、一单词或短语 ( 102 ) 地址、一级别 ( 103 ) 地址及一具体



信息的地址及长度 (104) 地址, 其中该索引编号 (101) 地址是用以存放每一单词或短语的索引编号, 该单词或短语 (102) 地址用以存放数个与过引编号 (101) 地址相对应的单词或短语, 该级别 (103) 地址用以存放数个与单词或短语 (102) 地址相对应的级别, 而该具体信息的地址及长度 (104) 地址则用以存放数个与单词或短语 (102) 地址相对应的单词或短语的位置及长度。例如: 索引编号 1 表示第一个词汇, 单词 A 表示具体的单词内容为 A, 级别 1 表示该单词 A 为初级, 而具体信息的地址及长度为 1,121 表示该单词 A 的具体信息的存放位置及长度, 该数据库根据以上所述的存放规则, 将词汇进行整理及排序, 以利快速查找(如图 2 所示)。

10 请参阅图 3 所示, 本发明由电脑的中央控制单元 20, 协调本发明整个系统的控制及动作, 该中央控制单元 20 分别与一存储器 40 及一定时器 50 相连接, 以相互传输信息, 该存储器 40 另分别与一数据存放装置 30 及一快取存储器 60 相连接, 以相互传输信号, 该快取存储器 60 与一显示器介面卡 70 相连接, 该显示器介面卡 70 并与一显示器 80 相连接, 以传输图像信号。

15 该数据存放装置 30 中存放有具体信息的地址及长度 (104) 的所有数据, 该存储器 40 中存放有级别 (103) 的所有数据, 而该快取记忆体 60 可通过中央控制单元 20 及定时器 50 的协调, 对所查找的文章进行分析及运算, 并将所得结果, 经显示器介面卡 70 显示在显示器 80 上。

请参阅图 4 所示, 本发明先将数据库读入, 以建立相对应的索引编号、

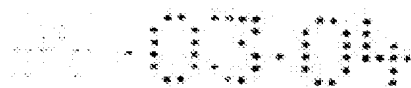
20 级别及具体信息的地址及长度的索引关系, 并将不同级别的单词及短语, 放入其相对应的存储空间内 (401), 再输入一篇文章并设定当前的搜索单词序号(Our-Word)为零 (402), 计算文章中有几句、几个单词 (403), 再判断计算的单词个数是否大于数据库内词汇总数 (404), 如是, 则结束, 如否, 则继续再判断句数标记是否小于数据库内总句数 (405), 如否, 则结束, 如是, 则继续通过一标志量(Flag), 用以区分当前所处理的句子是否在同一显示屏上 (406), 如在, 则置标志量(Flag) = false, 并继续进行显示 (407), 且回到 (404) 重复上述的步骤; 如不在当前的显示屏内, 则将当前的显示屏刷新, 重新计算当前的显示位置并显示其相对应的内容, 同时, 将搜索单词序号(Our-Word)加 1 并置标志量(Flag) = true, 且回到 (404)

25 重复上述的步骤。

30 重复上述的步骤。

请参阅图 4 和图 5 所示, 本发明的计算文章中有几句、几个单词 (403)





(如图 4 所示), 其根据使用者设定的单词级别 ( 201 ), 开始搜索显示器上英文文章的词汇, 并判断是否已搜索到文章尾 ( 202 ), 如是, 则将所搜索到的词汇, 进行 ( 404 ) 的操作; 如否, 则将该篇文章根据自然语言段落规则, 分割成数个独立的句子, 并于总句数加 1 ( 203 ), 以令再次继续进行搜索, 并再次判断是否已搜索到句尾 ( 204 ), 如是, 则进行 ( 202 ) 的操作, 如否, 则依规则取出其中的单词, 过滤重复的单词, 并将总句数加 1 ( 205 ), 再重复 ( 204 ) 的操作。

请参阅图 4, 图 6 和图 7 所示, 本发明的前标志量 (Flag) = false ( 407 ) 及标志量 (Flag) = true ( 408 ) (如图 4 所示), 其将根据所搜索到的词汇, 计算该词汇在显示器上所对应的位置 (行, 栏) 及长度 ( 206 ), 并重新调整该词汇所对应的行与相邻行之间距离, 为一般行与行之间距离的两倍 ( 207 ), 通过数据库的具体信息的地址及长度 ( 104 ), 及通过该词汇所对应的位置, 计算出该词汇的中文翻译存放的位置与长度 ( 208 ), 而将该词汇所对应的中文翻译正确放在相邻该词汇的相对应位置上 ( 209 ), 并于该中文翻译下方加一下划线 210, 令使用者可清楚阅读词汇的中文翻译 (如图 7 所示)。

在本发明中, 请参阅图 8 和图 9 所示, 该数据库并提供一界面 (如图 9 所示), 使使用者可针对本身的不同英文程度, 选择移动 ( 301 ), 以自由调整所有单词及短语的难、易级别, 使当某一级别的部分词汇, 对该使用者而言, 已变成较简单时, 可令使用者任意地降低该各词汇的级别, 或当某一级别的部分词汇, 对该使用者而言, 已变成较生熟时, 可令使用者任意地提升该各词汇的级别, 同时, 对于不在数据库中的词汇, 使用者也可选择加入 ( 302 ), 将不在数据库中的词汇, 存放数据库中的相对应级别中, 或选择删除 ( 303 ), 将欲删除的词汇, 从数据库的相对应级别中删除。

以上所述的, 仅为本发明最佳的一具体实施例, 但本发明的构造特征并不局限于此, 任何熟悉该项技术人员在本发明领域内, 可轻易地变化或修饰, 都可涵盖在所附的本发明权利要求范围。

# 说明书附图

---

索引编号	(101)
单词或短语	(102)
相应的级别	(103)
具体信息的存放位置及长度	(104)

图 1

索引编号(101)	单词或短语(102)	级别(103)	具体信息的存放位置及长度(104)
1	A	1	1,427
7680	California	1	12328,21
21211	As soon as	3	47829, 673

图 2

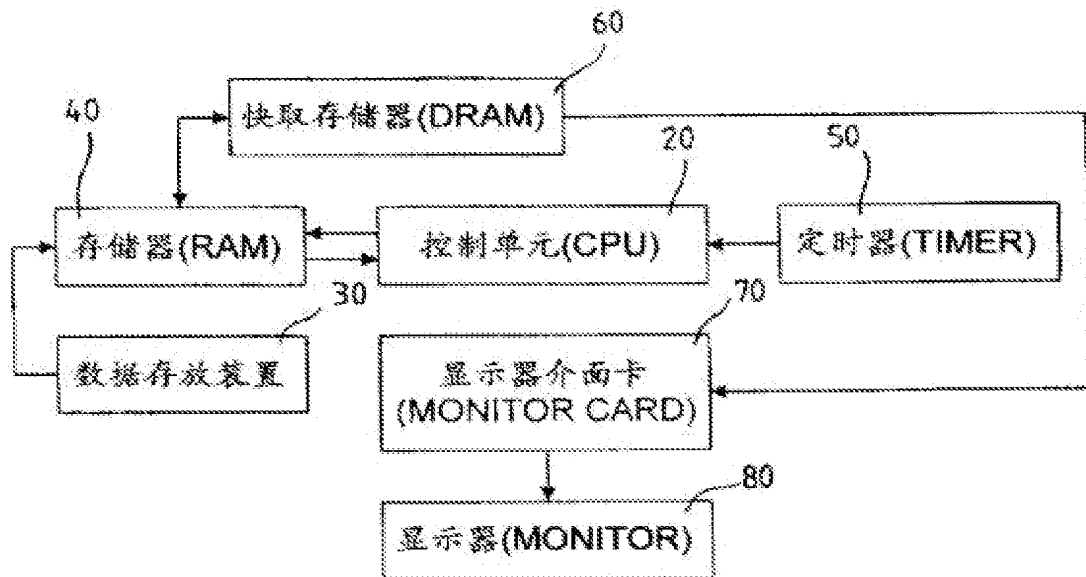


图 3

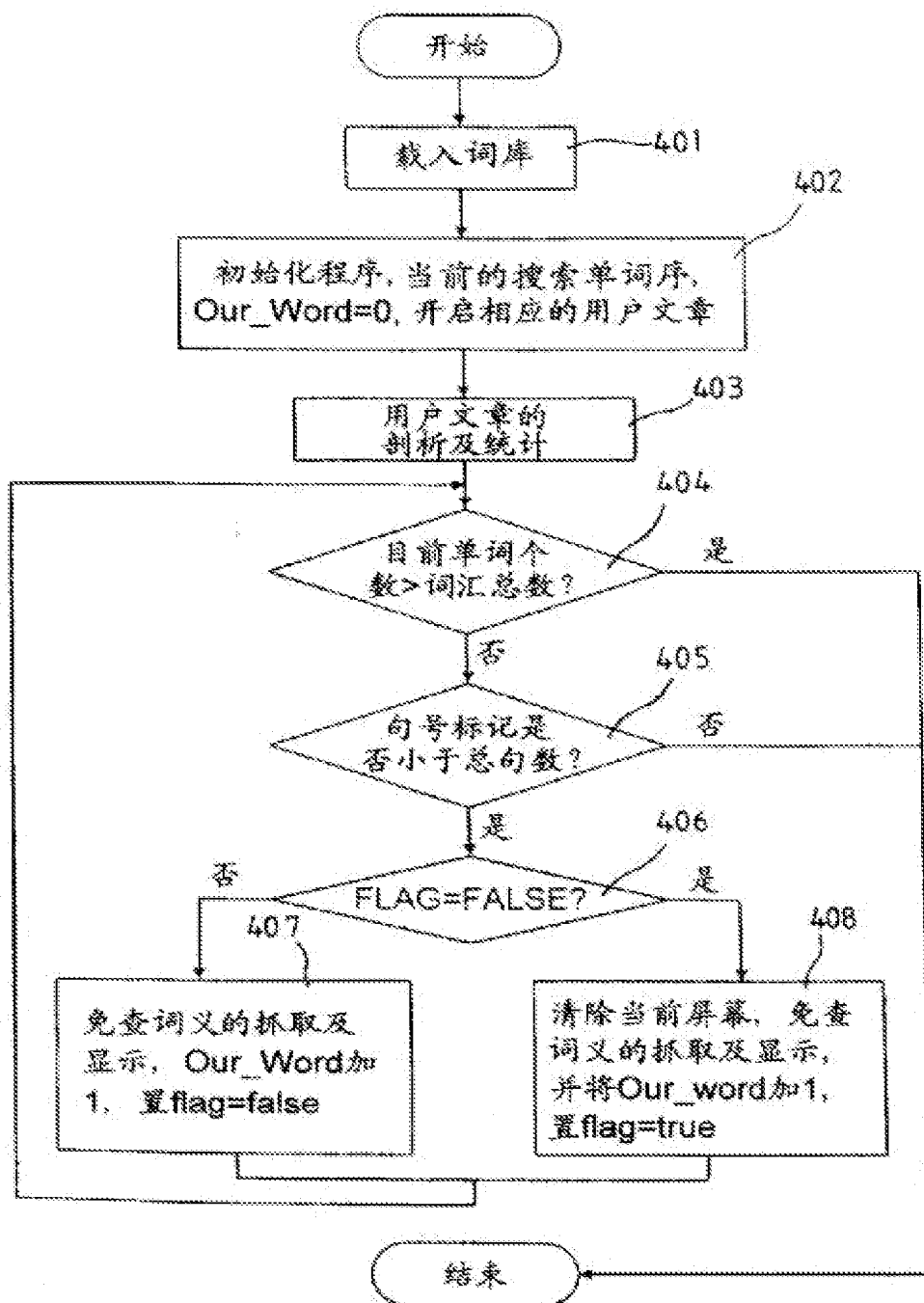


图 4

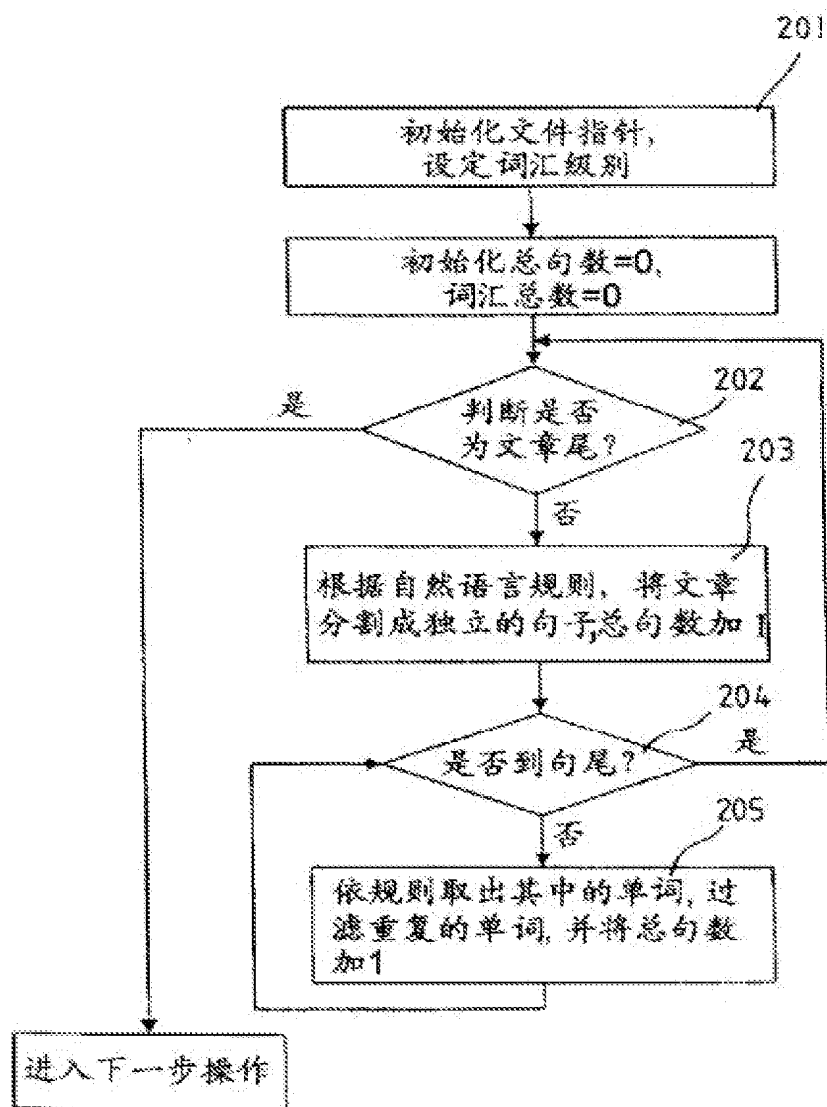


图 5

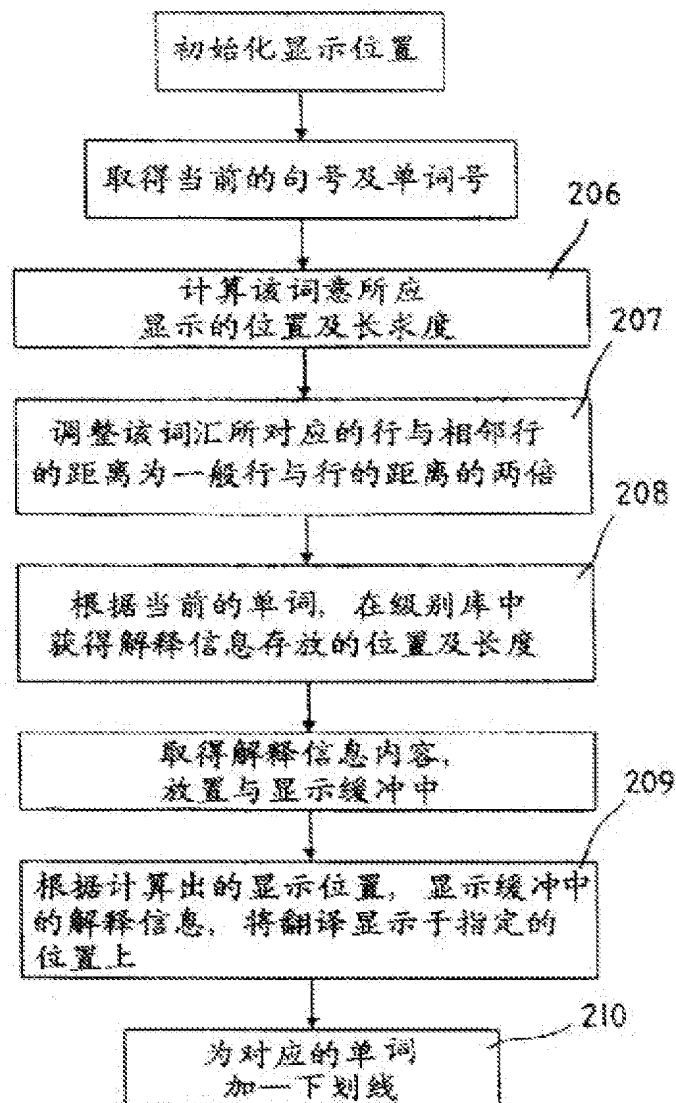


图 6

1990年10月10日  
1990年10月10日  
1990年10月10日

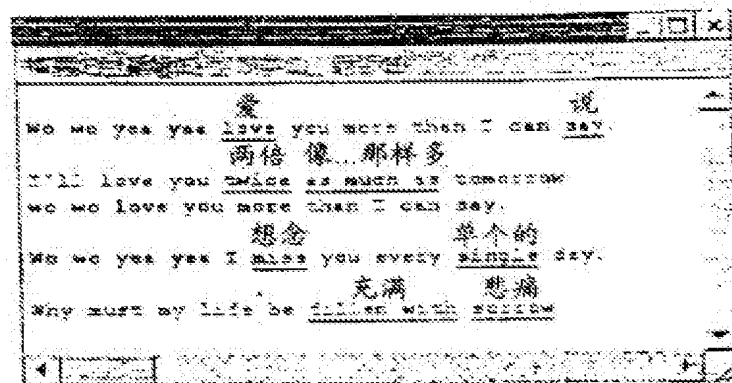


图 7



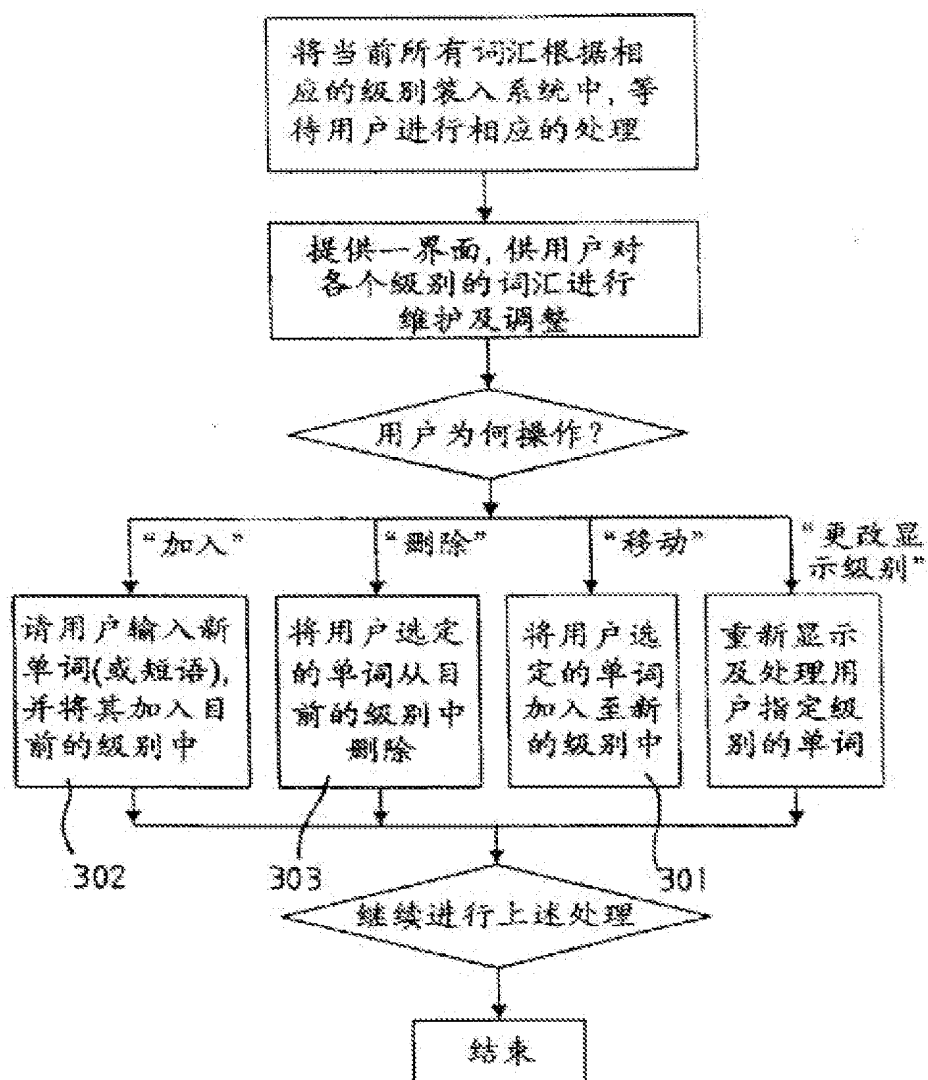


图 8

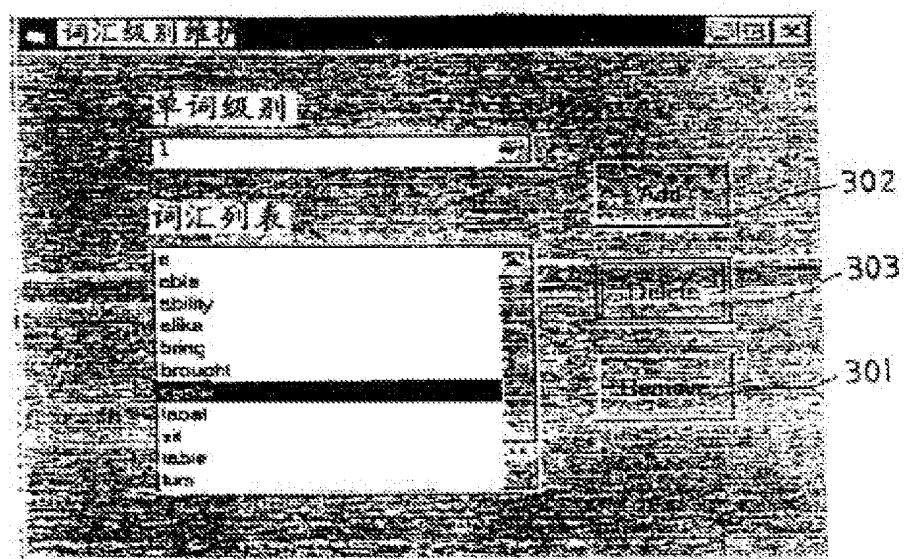


图 9

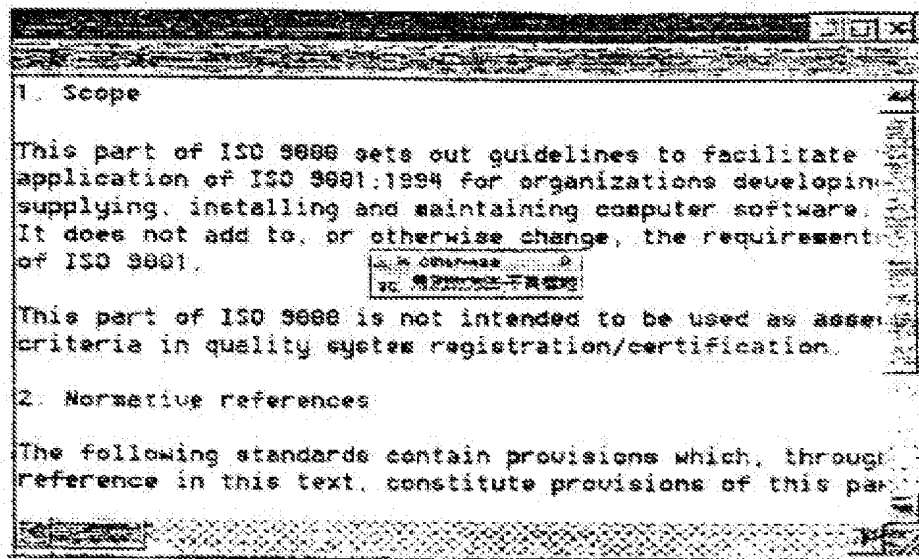


图 10